LIBRARY OF THE
UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

510.84
I l6r
no.61-80
cop.3

[redacted]

Digitized by the Internet Archive
in 2013

http://archive.org/details/fundamentalsofju64popp

UNIVERSITY OF ILLINOIS
GRADUATE COLLEGE
DIGITAL COMPUTER LABORATORY

INTERNAL REPORT NO. 64

FUNDAMENTALS OF JUNCTION TRANSISTOR PHYSICS

By

W. J. Poppelbaum

August 30, 1955

This work has been supported by
Contract N6ori-71 Task XXIV
United States Navy
ONR NR 048 094

TABLE OF CONTENTS

1. The Number of Possible States of a System of Particles 1

2. Number of Occupied States. Fermi Statistics 2

3. The Physical Meaning of the Fermi-level 4

4. Electrons and Holes. Energy Bands. Electron and Hole Densities 5

5. Origin of Energy. Intrinsic Densities. Equation for the Electrostatic Potential 7

6. P-regions and n-regions. ϕ and ψ in an Equilibrium Junction. Quasi Fermi-levels 9

7. Mobilities, Diffusion Constants, Einstein Relationship. Current Densities expressed as Gradients of Quasi Fermi-levels 11

8. Generation and Recombination of Holes and Electrons. Continuity Equation 13

9. Variation of ϕ_n, ϕ_p and ψ in a Non-equilibrium pn Junction 14

10. The Voltage-current Characteristic of a pn Junction 17

11. Variation of ϕ_n, ϕ_p and ψ in a Non-equilibrium pnp Transistor 19

12. The Voltage-current Characteristics of a pnp Transistor 22

13. Behavior of the Base Region at High Injection Levels 24

14. Carrier Delay in the Base Region as a Function of Injection Level 25

15. The Current Amplification α of a Transistor 28

16. The α Cutoff Frequency as a Function of Injection Level 31

17. The Transition Region Capacitance 33

18. The Collector Space Charge Layer 36

Acknowledgement

The deepest gratitude is due to Professor J. Bardeen for his helpful criticism and the suggestions he has made for the improvement of many sections.

INTRODUCTION

The following sections are an introduction to the theories used in the discussion of transistor behavior. To simplify explanations we only consider pnp transistors: all results can easily be transformed for npn units. The text is in no way original except for details of exposition and certain conclusions reached in the latter sections of the report. The exposition leans on diverse treatises on solid state physics and a number of articles which have become classic in the field. The principal sources used are:

J. Bardeen:	Fundamentals of Transistor Devices. Lectures at the Graduate Summer School in Semiconductors (1953).
R. C. Tolman:	The Principles of Statistical Mechanics. Oxford University Press (1950).
F. Seitz:	Modern Theory of Solids, McGraw-Hill (1940).
W. Shockley:	Electrons and Holes in Semiconductors, Van Nostrand (1950).
W. Shockley, M. Sparks, G. K. Teal:	PN Junction Transistors, Physical Review, vol. 83, No. 1, pp. 151-162 (July 1951).
W. Shockley:	The Theory of pn Junctions in Semi-conductors and pn Junction Transistors, BSTJ, vol. 28, pp. 435-489, (July 1949).
C. Herring:	Theory of Transient Phenomena in the Transport of Holes in an Excess Semiconductor, BSTJ, vol. 28, pp. 402-469 (July 1949).
E. S. Rittner:	Extension of the Theory of the Junction Transistor, Physical Review, vol. 94, No. 5, pp. 1161-1171 (June 1954).
T. Misawa:	Diffusion Capacitances and High Injection Level Operation of Junction Transistors, Proceedings I.R.E. p. 749 (June 1955).
J. M. Early:	Effects of Space-charge Layer Widening in Junction Transistors, Proceedings I.R.E., p. 1401 (November 1952).

1. The Number of Possible States of a System of Particles

Let V be a volume in free space containing N particles (for example electrons). Let $\vec{x}$ be the radius vector from an arbitrary origin O and let $\vec{\sigma} = m\vec{v}$ be the momentum vector of a particle. It is convenient to represent the succession of states of movement $(\vec{x},\vec{\sigma})$ in a 6 dimensional space called phase space and having coordinates $x_1 x_2 x_3$ $\sigma_1 \sigma_2 \sigma_3$. The fundamental assumptions of quantum mechanics can then be summarized as follows:

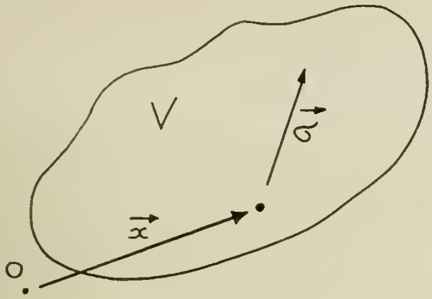

Figure 1

Coordinate Vector and Momentum Vector

> If by any apparatus we try to determine the components of $\vec{x}$ and $\vec{\sigma}$ simultaneously, we will encounter uncertainties Δx_i and $\Delta\sigma_i$ such that for every value of i = 1, 2, 3 $\Delta x_i \Delta\sigma_i \cong h$, where h is Planck's constant.
>
> (Heisenberg's Principle)

We deduce from Heisenberg's Principle that if we divide phase space into cells of size $\Delta x_1 \Delta x_2 \Delta x_3 \Delta\sigma_1 \Delta\sigma_2 \Delta\sigma_3 = h^3$, we cannot distinguish the state of movement of two particles in the same cell. Statistics applied to electrons are based upon the following hypothesis:

> Each cell of size h^3 in phase space can only accomodate 2 particles (of opposite magnetic momenta or "spins") at the most.
>
> (Pauli's Principle)

Remark: We can apply the above theory even in the case where the electrons are inside a crystal. Then the "crystal momentum" $\vec{\sigma}$ is no longer equal to $m\vec{v}$ but still equals $\frac{h}{2\pi}\vec{k}$ where $\vec{k}$ is the wave vector of the associated wave.

Notice that each cell corresponds to a definite energy $E = E(\vec{x},\vec{\sigma})$ but that for a given energy there may be several cells.

Let us now choose a volume $V = \Delta x_1 \Delta x_2 \Delta x_3$ small enough to neglect the variation of $\vec{x}$; then $E = E(\vec{\sigma})$ and the size of each cell in $\vec{\sigma}$-space ("momentum space") becomes

$$\Delta\sigma_1 \Delta\sigma_2 \Delta\sigma_3 = \frac{h^3}{V} .$$

This cell contains again at most 2 particles.

We can now calculate the number of possible states having energies between E and E + dE if we know E as a function of $\vec{\sigma}$. It turns out that very often

$$E = E_n + \frac{\sigma^2}{2m} \quad , \tag{1.1}$$

where E_n = potential energy of particle (zero of energy to be fixed later) ,

$\frac{\sigma^2}{2m}$ = kinetic energy.

Now the volume of a spherical shell between the radii σ and $\sigma + d\sigma$ is $4\pi\sigma^2 d\sigma$. The number of states $N(\sigma)d\sigma$ corresponding to this shell is therefore

$$N(\sigma)d\sigma = \frac{2V}{h^3} 4\pi\sigma^2 d\sigma . \tag{1.2}$$

Let E and E + dE be the energies corresponding to σ and $\sigma + d\sigma$ by (1): differentiating

$$dE = \frac{\sigma}{m} d\sigma. \tag{1.3}$$

Substituting (3) and (1) into (2) we obtain for the number of possible states N(E)dE between E and E + dE

$$N(E)dE = \frac{4\pi V}{h^3} (2m)^{3/2} (E - E_n)^{1/2} . \tag{1.4}$$

2. Number of Occupied States. Fermi Statistics.

The total number of possible states can be much larger than the number N of particles in V. We shall now try to determine the percentage of the possible states which are occupied. For this purpose we divide the states (corresponding to a small volume V) into groups 1, 2, ..., r, ... such that the r^{th} group has an energy between E_r and $E_r + dE_r$. Suppose that there are S_r possible states in the group, N_r of which are occupied. Let us count the number of ways in which the distribution $N_1 N_2 \ldots N_r \ldots$ can be set up: This number is called the probability W of the distribution (notice that W is not normalized). By the laws of permutations

$$W = \prod_r \frac{S_r!}{N_r!(S_r - N_r)!} \tag{2.1}$$

It is evident that the physically realized distribution will render W maximum: small perturbations satisfying

$$\Sigma N_r = \text{Constant } N \text{ (constant number of particles)} \qquad (2.2)$$

$$\Sigma E_r N_r = \text{Constant } E \text{ (constant total energy)} \qquad (2.3)$$

must therefore leave W (or lnW) unchanged. Let 1 and 2 be two perturbations such that 1 changes E by δE_1 and lnW by $(\delta \ln W)_1$ while 2 changes E by δE_2 and lnW by $(\delta \ln W)_2$. Then

$$(\delta \ln W)_1 + (\delta \ln W)_2 = \delta \ln W = 0$$

and $$\delta E_1 + \delta E_2 = 0 \qquad \text{(by 2.3)}$$

simultaneously. Since the perturbations are arbitrary, this means that

$$\delta \ln W = \beta \delta E \qquad (2.4)$$

where β is a quantity independent of E and W. By comparing the result of statistical calculations to classical thermodynamics, it can be shown that

$$\beta = \frac{1}{kT} \qquad (2.5)$$

where k is Boltzmann's constant and T the absolute temperature.

Now consider a perturbation more in detail. Suppose it consists in the transfer of δN_r particles from group r to group s. Then $\delta N_s = -\delta N_r$ and $\delta E = (E_r - E_s)\delta N_r$. Further (2.1) gives

$$\delta \ln W = -\delta \ln(N_r!) - \delta \ln[(S_r - N_r)!]$$
$$-\delta \ln(N_s!) - \delta \ln[(S_s - N_s)!].$$

Since $\ln x! \overset{\sim}{=} x \ln x$, we have

$$\delta \ln W = -\delta N_r \ln N_r + \delta N_r \ln(S_r - N_r) - \delta N_s \ln N_s + \delta N_s \ln(S_s - N_s)$$

$\delta(S_r - N_r) = -\delta N_r$ etc.!) and by using (2.4) and (2.5)

$$\frac{1}{kT} = \frac{\delta \ln W}{\delta E} = \frac{1}{E_r - E_s} [-\ln N_r + \ln(S_r - N_r) + \ln N_s - \ln(S_s - N_s)]$$

which gives

$$\frac{E_r}{kT} + \ln \frac{N_r}{S_r - N_r} = \frac{E_s}{kT} + \ln \frac{N_s}{S_s - N_s} \quad . \qquad (2.6)$$

This being true for all groups r and s, we shall set each side equal to E_F/kT and call E_F the Fermi-level. Then for any r

$$\ln \frac{N_r}{S_r - N_r} = (E_F - E_r)/kT. \tag{2.7}$$

Now introduce the distribution function (value for group r) by

$$f_r = \frac{N_r}{S_r} . \tag{2.8}$$

Then

$$f_r = \frac{1}{1 + e^{(E_r - E_F)/kT}}$$

i.e., for any given energy range E, E + dE the quotient (possible states)/(occupied states) is

$$f = \frac{1}{1 + e^{(E-E_F)/kT}} . \tag{2.9}$$

3. The Physical Meaning of the Fermi-Level

As can be seen from (2.9) $f = 1/2$ when $E = E_F$, i.e. 50% of the states are occupied for an energy equal to the Fermi-level.

We are now going to prove that if we transfer δN electrons (of charge-q) from a conducter in equilibrium at level E_{F1} and electrostatic potential V_1 to a conductor 2 in equilibrium at level E_{F2} and potential V_2 in a reversible fashion,

$$-q(V_1 - V_2) = E_{F1} - E_{F2} \tag{3.1}$$

or, for any origin of E_F and V

$$V = -\frac{E_F}{q} + \text{constant} \tag{3.2}$$

It is possible to show by irreversible thermodynamics that in the most general case of flow of electric and thermal currents (3.2) is still true.

To prove (3.1), we transfer δN electrons from conductor 1 to a conductor 2 at the same temperature in a reversible fashion (total $\delta S = 0$ and therefore $\delta F = \delta E - T\delta S = \delta E$, where F is the free energy). Now $\delta F = \delta F_1 + \delta F_2$ and by a well-known theorem of thermodynamics S = k lnW (in the union of 2 systems we have to add entropies and multiply probabilities!). Therefore

$$\delta F_2 = \delta E_2 - T\delta S_2 = \delta E_2 - kT(\delta \ln W)_2 \tag{3.3}$$

Suppose that we add the electrons to states in group r in conductor 2. Then by the preceding section

$$(\delta \ln W)_2 = -\ln\left(\frac{N_r}{S_r - N_r}\right)_2 \delta N$$

It follows that

$$\begin{aligned} \delta F_2 &= \delta E_2 + kT\,\delta N \ln\left(\frac{N_r}{S_r - N_r}\right)_2 \\ &= E_{F2}\,\delta N \end{aligned} \tag{3.4}$$

by utilizing (2.6). We see that δF_2 is independent of the group of states to which the electrons are added.

Using $\delta F = \delta F_1 + \delta F_2$ we find that

$$\delta F = (E_{F2} - E_{F1})\delta N \tag{3.5}$$

while

$$\delta E = q(V_1 - V_2)\delta N \tag{3.6}$$

As mentioned above, we have in this particular case $\delta F = \delta E$. Equation (3.1) thus follows immediately.

4. Electrons and Holes. Energy Bands. Electron and Hole Density

By a more detailed examination of the quantum mechanics involved in the movement of electrons inside a crystal, it can be shown that these electrons can only have energies in certain permitted ranges called bands. In the most common case there is a nearly empty band in which the electrons behave more or less like free particles: this is the so-called conduction band. Then there is a nearly filled band: the valency band. Electrons in the valency band are better described by the behavior of the vacant energy levels or holes. Practically a hole behaves like a particle of charge +q having approximately the electronic mass m.

We shall now calculate the electron density in the conduction band. For this we set $V = 1$ in (1.4), obtaining the number of possible states per unit volume:

$$N(E)dE = \frac{4\pi}{h^3}(2m)^{3/2}(E - E_n)^{1/2}$$

The fraction f by which we must multiply to find the occupied states is given by (2.9):

$$f = \frac{1}{1 + e^{(E-E_F)/kT}}$$

The number of occupied states between the energy limits E_1 and E_2 of the conduction band is thus

$$n = \int_{E_1}^{E_2} \frac{N(E)dE}{1 + e^{(E-E_F)/kT}} \tag{4.1}$$

Now this is obviously the electron density. We shall calculate (4.1) under the hypothesis that $E > > E_F$, then

$$\frac{1}{1 + e^{(E-E_F)/kT}} = e^{-(E-E_F)/kT} \tag{4.2}$$

and

$$n = e^{E_F/kT}\,\frac{4\pi(2m)^{1/2}}{h^3}\int_{E_1}^{E_2}(E - E_n)^{1/2}\,e^{-E/kT}dE$$

Putting $E^* = E - E_n$ and taking account of the rapid convergence (which permits to extend the upper limit to $+\infty$), we may write

$$\begin{aligned} n &= e^{(E_F-E_n)/kT}\,\frac{4\pi(2m)^{3/2}}{h^3}\int_0^{\infty} e^{-E^*/kT}dE^* \\ &= N_n\,e^{-(E_n-E_F)/kT} \end{aligned} \tag{4.3}$$

where

$$N_n = 2(2\pi mkT/h^2)^{3/2} = \text{"effective density of electrons in the conduction band."} \qquad (4.4)$$

By reasoning on the holes in the valency band, the potential energy E_p being necessarily measured from the same origin as E_n, we find for the hole density

$$p = N_p \, e^{-(E_F - E_p)/kT} \qquad (4.5)$$

where

N_p = "effective density of holes in the valency band."

5. Origin of Energy. Intrinsic Densities. Equation for Electrostatic Potential

We shall now proceed to choose the origin of the energies E, E_n, E_p, E_F in such a way that there will be some simplification in the formalism. Let

ψ = electrostatic potential (measured from infinity)

E_G = energy gap between the top of the valency band and the bottom of the conduction band.

The energy origin will then be chosen in such a way that

$$E_n = -q\psi + \frac{1}{2}\,kT\,\ln\left(\frac{N_n}{N_p}\right) + \frac{E_G}{2} \qquad (5.1)$$

leading immediately to

$$E_p = +q\psi + \frac{1}{2}\,kT\,\ln\left(\frac{N_n}{N_p}\right) - \frac{E_G}{2}\ . \qquad (5.3)$$

Further let us introduce a Fermi-potential ϕ such that

$$E_F = -q\phi \qquad (5.3)$$

Then (4.3) and (4.5) can be written

$$\begin{aligned} n &= n_i \, e^{q(\psi-\phi)/kT} \\ p &= n_i \, e^{q(\phi-\psi)/kT} \end{aligned} \qquad (5.4)$$

where

$$n_i = \sqrt{N_n N_p} \; e^{-E_G/kT} = \sqrt{pn} \; . \qquad (5.5)$$

It can be seen that for $\psi = \phi$, $p = n = n_i$: n_i is called the intrinsic carrier density. (4.4) shows that n_i is independent of the location in the crystal as long as E_G is constant: This is usually assumed to be the case. Formulae (5.4) are therefore perfectly general, n_i having the same value everywhere; only ϕ and ψ vary with the location.

We can now examine the case in which the crystal contains electrons, holes and ionized atoms. Let N_d be the density of positively ionized atoms (called donors) and N_a the density of negatively ionized atoms (called acceptors). The distribution of donors and acceptors is determined during the growing of the crystal. We thus obtain for the charge density

$$\rho = q(p - n + N_d - N_a) . \qquad (5.6)$$

Substituting for n and p the values (5.4) we have

$$\rho = q[N_d(\vec{x}) - N_a(\vec{x}) + 2n_i \sinh q\{\phi(\vec{x}) - \psi(\vec{x})\}/kT] . \qquad (5.7)$$

If we specify $\phi(\vec{x})$ we can calculate ψ from Poisson's equation

$$-\mathrm{div}\, E = \mathrm{lap}\, \psi = - \frac{\rho}{\epsilon_r \epsilon_o} \; {}^{*)} \qquad (5.8)$$

(MKS-system, ϵ_o = dielectric constant of vacuum
ϵ_r = relative dielectric constant of crystal)

In the treatment of transistor problems two cases are usually considered: as a first approximation a zero space charge solution ($\rho = 0$) at low current densities and then a zero space charge solution at high current densities. Only in exceptional cases is one able to treat the case $\rho \neq 0$: we shall not attempt to do so until section 18.

---

*) In the one dimensional case lap $= \dfrac{\partial^2 \psi}{\partial x^2}$

6. p-regions and n-regions. ϕ and ψ in an Equilibrium Junction. Quasi Fermi-levels

p		n	
N_{ap}	N_{dp}	N_{an}	N_{dn}
p_p	n_p	p_n	n_n

Figure 2

Notation for p-regions and n-regions

Consider two adjacent regions distinguished by the indices p and n such that in the left hand region and in the right hand region respectively we have

$$\left.\begin{aligned} N_{ap} - N_{dp} &= p_o > 0 \\ N_{an} - N_{dn} &= -n_o < 0 \end{aligned}\right\} \qquad (6.1)$$

Let us consider the zero space charge case; then (5.6) gives

$$\left.\begin{aligned} p_p &= p_o + n_p \\ n_n &= n_o + p_n \end{aligned}\right\} \qquad (6.2)$$

We shall suppose that

$$\left.\begin{aligned} p_p &>> n_p \\ n_n &>> p_n \end{aligned}\right\} \qquad (6.3)$$

The p-region thenpractically only contains holes, while the n-region practically only contains electrons. Electrons in the p-region or holes in the n-region are called minority carriers.

Let us try to gain some insight into the equilibrium behavior of ϕ and ψ when we go from a p-region into an n-region across a "pn junction" i.e. the transition region. The important thing to notice is that ϕ must have the same value throughout the sample: if it were not so, energy changes according to section 3 would accompany the transfer of electrons (or holes) from one region to the other and this contradicts the hypothesis of equilibrium. This does not imply by any means that ψ is the same everywhere. But we can see that ψ must be constant at all locations where the concentration of carriers $p_p n_p$ and $p_n n_n$ is constant, i.e. inside the p-region and inside the n-region. This is evident because carriers can only drift as a consequence of concentration gradients and electric fields $E = -\text{grad}\,\psi$*): if there is equilibrium and no concentration gradient, there can be no potential gradients either.

---

*) In the one dimensional case $\text{grad}\,\psi = \frac{\partial \psi}{\partial x}$

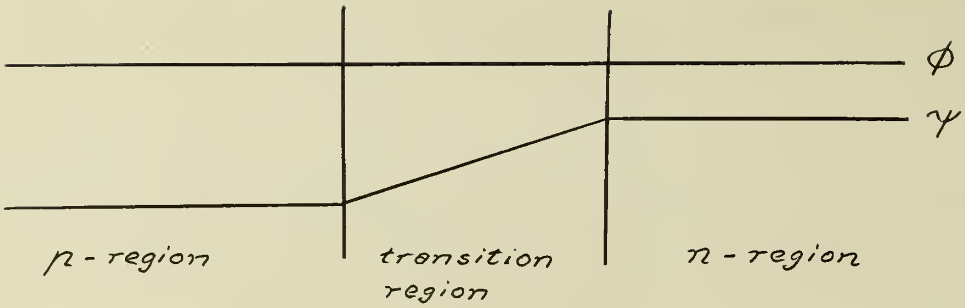

Figure 3

ϕ and ψ in an Equilibrium pn Junction

In the transistion zone however a big potential gradient must be present to offset the influence of the rapidly varying concentration of holes and electrons: ϕ and ψ behave as indicated in Figure 3.

In equilibrium the computation of p_n and n_p (called p_n^o and n_p^o to indicate equilibrium!) offers no difficulty when p_o and n_o are given. (5.4) and (5.5) are valid, therefore $(p_o + n_p^o)n_p^o = n_i^2$ and $(n_o + p_n^o)p_n^o = n_i^2$, giving by (6.3)

$$\left.\begin{aligned} n_p^o &= n_i^2/p_o \\ p_n^o &= n_i^2/n_o \end{aligned}\right\} . \qquad (6.4)$$

Let us now take the non-equilibrium case. Then we do not know anything a priori about the Fermi-levels in the p- and n-regions. Worse still: when current flows across a pn junction we are evidently no longer entitled to use formulae (5.4) which are a consequence of (4.3) and (4.5) and therefore valid only under equilibrium conditions.

Nevertheless we can write relationships analogous to (5.4) containing "quasi Fermi-levels" (sometimes called "imrefs") ϕ_n and ϕ_p to be defined in a suitable way:

$$\begin{aligned} n &= n_i e^{q(\psi-\phi_n)/kT} \\ p &= n_i e^{q(\phi_p-\psi)/kT} \end{aligned} \qquad (6.5)$$

Solving (6.4) for ϕ_n and ϕ_p we then obtain formulae defining the quasi Fermi-levels:

$$\phi_n = \psi - \frac{kT}{q} \ln\left(\frac{n}{n_i}\right),$$
$$\phi_p = \psi + \frac{kT}{q} \ln\left(\frac{p}{n_i}\right). \qquad (6.6)$$

In section 9 we shall look into the behavior of ψ, ϕ_n and ϕ_p in a junction when current flows across it, i.e. the behavior in the non-equilibrium case.

7. Mobilities, Diffusion Constants, Einstein Relationship and Current Densities Expressed as Gradients of Quasi Fermi-levels

The movement of electrically charged particles can come about in two ways: firstly they can attain a certain average speed due to the electric field present, the speed being limited because of collisions with the crystal lattice. It turns out that often the average speed is proportional to the electric field: one can then define a constant, called mobility μ, which is the average speed per unit electric field.

Secondly the movement can be caused by concentration gradients: the (random-) thermal velocity then produces a general drift to the regions of lower carrier density. The number of carriers drifting across unit area in unit time often turns out to be proportional to the concentration gradient: one can then define a constant, called diffusion constant D, which is the particle current density per unit concentration gradient.

Take for example holes having a mobility μ_p and a diffusion constant D_p. Let p be their volume density and q their charge as before. Call $E = -\text{grad}\,\psi = -\frac{d\psi}{dx}$ the electric field. Then the total particle current density due to E and grad p is $\mu_p pE - D_p \text{ grad } p$, i.e. the current density I_p for holes is given by

$$I_p = q(\mu_p pE - D_p \text{ grad } p). \qquad (7.1)$$

In the same way we would have for electrons

$$I_n = q(\mu_n nE + D_n \text{ grad } n). \qquad (7.2)$$

Let us take an inhomogeneous p-region in equilibrium. Then we can write (5.4)

$$p = Ae^{-q\psi/kT} \qquad (7.3)$$

where $A = n_i e^{q\phi}$ has the same value throughout the region as discussed in section 6 but $\psi = \psi(x)$. Using (7.3) we obtain

$$\mathrm{grad}\ p = \frac{dp}{dx} = -p \frac{q}{kT} \frac{d\psi}{dx} = \frac{pq}{kT} E$$

Since we have equilibrium, $I_p = 0$ and (7.1) gives

$$0 = pqE(\mu_p - D_p \frac{q}{kT})$$

or

$$\mu_p = \frac{q}{kT} D_p \quad \text{(Einstein relationship)} \qquad (7.4)$$

In a similar way we can prove that

$$\mu_n = \frac{q}{kT} D_n \qquad (7.5)$$

We can now establish a relationship between the gradients of ϕ_p and ϕ_n and I_p and I_n in the general non-equilibrium case. We then have by equations (6.5)

$$\begin{aligned} n &= n_i e^{q(\psi - \phi_n)/kT} \\ p &= n_i r^{q(\phi_p - \psi)/kT} \end{aligned} \qquad (7.6)$$

Let us now use (7.1) and (7.2), remembering (7.4) and (7.5). Calculating* grad p from (7.6) we have

$$\mathrm{grad}\ p = p \frac{q}{kT} (\mathrm{grad}\ \phi_p - \mathrm{grad}\,\psi),$$

so that (7.1) gives

$$\begin{aligned} I_p &= q\mu_p pE - q \frac{kT}{q} \mu_p p \frac{q}{kT} (\mathrm{grad}\ \phi_p + E) \\ &= -q\mu_p p\ \mathrm{grad}\ \phi_p \end{aligned} \qquad (7.7)$$

In the same way we can prove that

$$I_n = -q\mu_p n\ \mathrm{grad}\ \phi_n \qquad (7.8)$$

---

*) It is sufficient to consider the one-dimensional case in which grad $p = \frac{\partial p}{\partial x}$.

8. Generation and Recombination of Holes and Electrons. Continuity Equation

If we isolate unit volume inside a semiconductor (not necessarily in equilibrium), the number of holes (or electrons) in this volume varies for several causes. First of all holes and electrons recombine (i.e. electrons "fall" into the vacant energy levels called holes) at a rate r particle pairs/unit time. Then holes are generated due to thermal excitation, light, etc. at a rate g particles/unit time. Let us try to estimate the rate of change g-r.

Take for instance holes in the n-region; then for equilibrium $p_n = p_n^o$ given by (6.4). Let g^o and r^o be the values of g and r in the equilibrium case: $g^o = r^o$. Thus $g-r = 0$ when $p_n^o - p_n = 0$. We shall now make the fundamental assumption that g is constantly equal to g^o, i.e. that the minority carrier concentration does not affect the creation of holes. Since the rate of recombination r must be proportional to p_n (for the probability of a recombination is proportional to p_n) we must have

$$g - r = g_o - \lambda p_n$$

Taking the equilibrium case we find that $g_o = \lambda p_n^o$. Further we deduce from the definition that λ has the dimensions $(\text{time})^{-1}$: we shall therefore introduce a constant $\tau_p = \lambda^{-1}$ called lifetime of holes in n-type material and write

$$(g - r)_{\text{holes in n-region}} = \frac{p_n^o - p_n}{\tau_p} . \tag{8.1}$$

In the same way we can treat electrons in p-type material by introducing a lifetime τ_n and writing

$$(g - r)_{\text{electrons in p-region}} = \frac{n_p^o - n_p}{\tau_n} . \tag{8.2}$$

Notice that (8.1) and (8.2) have been established for the non-equilibrium case, i.e. the case where current flows. We shall now express that in this case the change of the number of holes (or electrons) per unit volume and unit time is due to g-r and the holes carried into the volume by the current: this is the continuity equation.

Take again the case of holes in an n-region: I_p, the current density of holes, is given by (7.1). The total charge of the holes going out of unit volume is given by the divergence* of I_p; the number of holes going into unit volume in unit time is thus $-\frac{1}{q}\operatorname{div} I_p$. The continuity equation thus becomes

---

*) In the one dimensional case $\operatorname{div} I_p = \dfrac{\partial I_p}{\partial x}$.

$$\frac{\partial p_n}{\partial \tau} = \frac{p_n^o - p_n}{\tau_p} - \frac{1}{q} \operatorname{div} I_p \tag{8.3}$$

In exactly the same way we have for electrons in the p-region

$$\frac{\partial n_p}{\partial \tau} = \frac{n_p^o - n_p}{\tau_n} + \frac{1}{q} \operatorname{div} I_n \tag{8.4}$$

Remark: The above reasoning can be extended to holes in the p-region and electrons in the n-region by suitably redefining the τ. Calling $p^o (= p_n^o$ or $p_o)$ and $n^o (= n_p^o$ or $n_o)$ the equilibrium concentrations we have the general equations

$$\frac{\partial p}{\partial \tau} = \frac{p^o - p}{\tau_p} - \frac{1}{q} \operatorname{div} I_p \tag{8.5}$$

$$\frac{\partial n}{\partial \tau} = \frac{n^o - n}{\tau_n} + \frac{1}{q} \operatorname{div} I_n \tag{8.6}$$

9. Variation of ϕ_n, ϕ_p and ψ in a Non-equilibriun pn Junction

Let us go back to the problem of section 6, but supposing this time that the equilibrium is disturbed by modifying ψ at the left hand side of the p-region by $\delta\psi$, leaving ψ at the right hand side of the n-region unchanged.

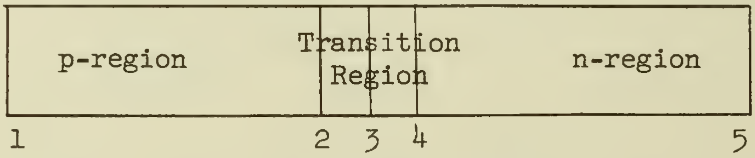

Figure 4

Significant Points in a pn Junction

We shall make 3 hypothesis:

(1) We shall assume that far from the junction, in 1 and 5, equilibrium conditions obtain. The carrier densitites will then be given by (6.4): $p(1) = p_o$, $n(1) = n_p^o$, etc. This means that almost the total current density I is produced by holes alone in 1 and by electrons alone in 5:

$$I = I_p(1) = I_n(5) \tag{9.1}$$

(2) We shall assume that the transition region 2,4 is so narrow that we can neglect recombination in it:

$$\left.\begin{aligned} I_p(2) &= I_p(4) \\ I_n(2) &= I_n(4) \end{aligned}\right\} \qquad (9.2)$$

(3) We shall assume that n_o and $p_o >> p_n$ and n_p: this is the so-called "low" injection level case." We shall partially abandon this hypothesis in section 13.

Using (7.7) we find that

$$|\text{grad } \phi_p| \leq \frac{I}{\mu_p p_o q} \qquad \text{in the p-region}$$

$$|\text{grad } \phi_n| \leq \frac{I}{\mu_p p_n q} \qquad \text{in the n-region}$$

Since $p_o >> p_n$ by hypothesis 3 ,

$$|\text{grad } \phi_p|_p << |\text{grad } \phi_p|_n$$

and it is reasonable to put

$$|\text{grad } \phi_p|_p = 0 \qquad (9.3)$$

i.e. ϕ_p is constant in the p-region.

In the same way we can find that

$$|\text{grad } \phi_n|_n = 0 \qquad (9.4)$$

i.e. ϕ_n is constant in the n-region.

As regards the transition region, we have to distinguish two cases. For forward and low reverse bias the carrier densities of either type will be sufficient to apply simultaneously (9.3) and (9.4). For high reverse bias there will be a very low carrier density and then ϕ_p and ϕ_n will vary in the transition region. Figure 5 gives the shape of ϕ_p and ϕ_n in the low bias case. Notice that the variation of ϕ_n in the p-region and that of ϕ_p in the n-region is arbitrary: it has no influence whatsoever on the other reasonings in the present section. By hypothesis (1) ϕ_p and ϕ_n must of course coincide at points 1 and 5: going back to (6.4) $p(1) = \frac{n_i^2}{p_o}$ and $p(1) = p_o$. Using (6.6) we obtain the announced result.

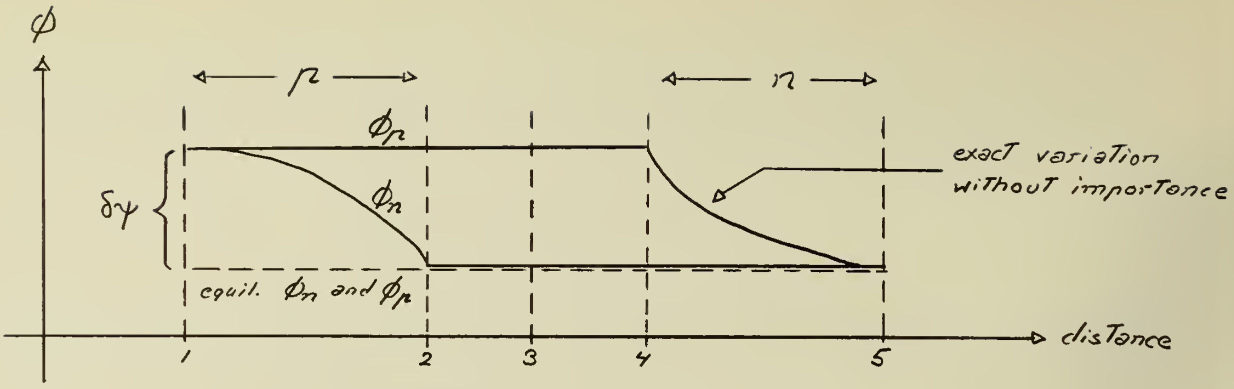
ϕ
p
n
ϕ_p
ϕ_n
$\delta\psi$
exact variation
without importance
equil. ϕ_n and ϕ_p
1
2
3
4
5
distance

Figure 5

ϕ_p and ϕ_n in a Non-equilibrium Junction

Let us now turn our attention to ψ. First of all (6.6) shows that ψ is constant between 1 and 2, for ϕ_p is constant and $p \sim p_o$ everywhere. In the same way we can prove that ψ is constant between 4 and 5. The variation of ψ is therefore as represented in Figure 6:

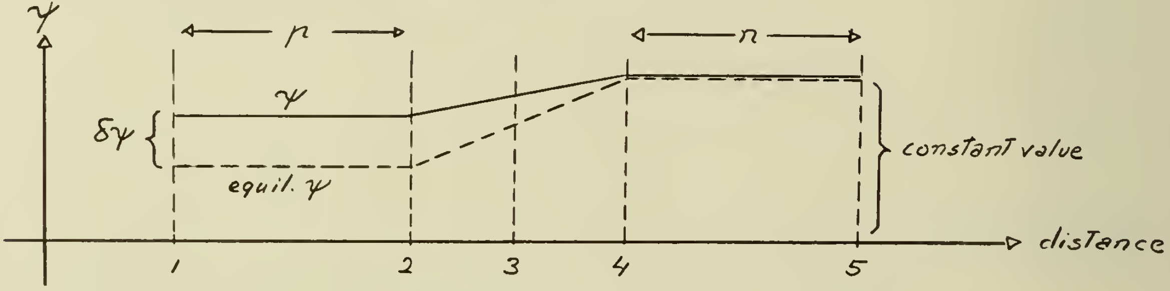
ψ
p
n
ψ
$\delta\psi$
equil. ψ
constant value
1
2
3
4
5
distance

Figure 6

ψ in a Non-equilibrium Junction

Lastly we can see that $\phi_p(1) - \phi_p(5) = \phi_n(1) - \phi_n(5) = \delta\psi$. The reason is again (6.6): $\phi_p - \psi = \frac{kT}{q} \ln \frac{p_o}{n_i}$ at point 1 whether we consider equilibrium or not. If we therefore increase ψ by $\delta\psi$, ϕ_p (and $\phi_n = \phi_p$) will increase by $\delta\psi$. The same reasoning applied to point 5 shows that here ϕ_p and ϕ_n keep their initial equilibrium case value. Summing up Figure 7 gives ϕ_p^o, ϕ_n^o and ψ^o for the equilibrium case in dashed lines while the heavy lines refer to a biased junction with ϕ_p, ϕ_n and ψ:

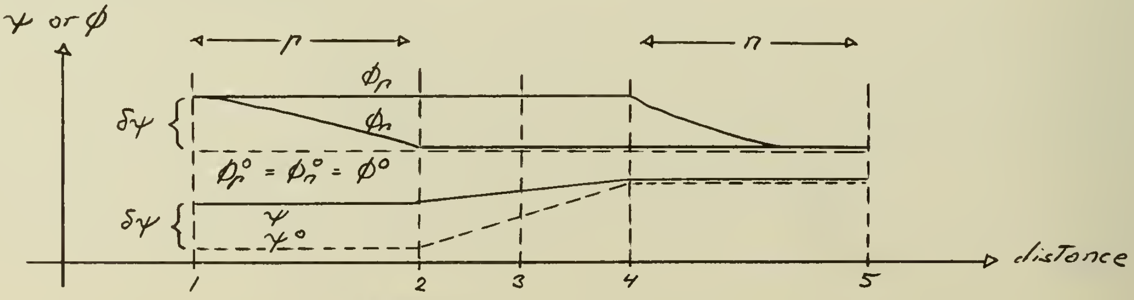
ψ or ϕ
p
n
ϕ_p
ϕ_n
$\delta\psi$
$\phi_p^o = \phi_n^o = \phi^o$
$\delta\psi$
ψ
ψ^o
1
2
3
4
5
distance

Figure 7

ψ, ϕ_p and ϕ_n in a Non-equilibrium Junction

10. The Voltage Current Characteristic of a p-n Junction

Let us again consider the junction of section 9. In point 4 (of coordinates x_4) we have from (6.5)

$$p_n(4) = n_i e^{q[\phi_p(4)-\gamma(4)]/kT} \tag{10.1}$$

$$n_n(4) = n_i e^{q[\gamma(4)-\phi_n(4)]/kT} \tag{10.2}$$

where p and n carry an index n to indicate the fact that we are in an n-region. Now by section 6 $n_n = n_o = (N_d - N_s)$ in an n-region because we neglect the space charge. (10.2) therefore gives

$$\frac{n_i}{n_o} \; e^{-q\phi_n(4)/kT} = e^{-q\gamma(4)/kT} .$$

Substituting into (10.1) we find that

$$p_n(4) = \frac{n_i^2}{n_o} \; e^{[\phi_p(4)-\phi_n(4)]/kT}$$

But by Figure 7 $\phi_p(4) - \phi_n(4) = \delta\gamma$ and by (6.9) $n_i^2/n_o = p_n^o$ = normal density of minority carriers in the n-region. The last equation therefore becomes

$$p_n(4) = p_n^o \; e^{q\delta\gamma/kT} . \tag{10.3}$$

Let us now try to determine $p_n(x)$ in the static case ($\frac{\partial p_n}{\partial \tau} = 0$) for $x \geq x_4$. The continuity equation (8.3) then becomes

$$0 = \frac{p_n^o - p_n(x)}{\tau_p} - \frac{1}{q} \operatorname{div} I_p = \frac{p_n^o - p_n(x)}{\tau_p} - \frac{1}{q} \frac{\partial I_p}{\partial x} . \tag{10.4}$$

To this we must join the equation giving I_p for a given grad $p_n(x)$ and a given E: γ being constant to the right of 4, E = 0 and (7.1) gives

$$I_p = -qD_p \operatorname{grad} p_n(x) = -qD_p \frac{\partial p_n(x)}{\partial x} \tag{10.5}$$

Inserting (10.5) into (10.4)

$$0 = \frac{p_n^o - p_n(x)}{\tau_p} + D_p \frac{\partial^2 p_n(x)}{\partial x^2} \tag{10.6}$$

P being a constant, we can write the solution to (10.6) which satisfies the boundary condition p_n finite for $x \to \infty$ in the form

$$p_n = p_n^o + Pe^{(x_4 - x)/Lp} \tag{10.7}$$

where
$$L_p = \sqrt{D_p \tau_p} \tag{10.8}$$

is called the diffusion length of holes in n-type material.

P can now be evaluated by the second boundary condition (10.3)

$$p_n^o \, e^{q\delta\psi/kT} = p_n(4) = p_n^o + P \quad \text{for } x = x_4$$

giving

$$p_n(x) = p_n^o + p_n^o(e^{q\delta\psi/kT} - 1) \, e^{(x_4 - x)/Lp} \tag{10.9}$$

As a last step we evaluate $I_p = -qD_p \dfrac{\partial p_n}{\partial x}$ at point 4. (10.9) gives directly

$$I_p(4) = I_{ps}(e^{q\delta\psi/kT} - 1) \tag{10.10}$$

where

$$I_{ps} = \frac{qp_n^o D_p}{L_p} \tag{10.11}$$

is the saturation hole current density in the junction. We see that for $\delta\psi$ big and negative $I_p = -I_{ps}$.

In a similar way we could discuss the electron current at point 2:

$$I_n(2) = I_{ns}(e^{q\delta\psi/kT} - 1) \tag{10.12}$$

where

$$I_{ns} = \frac{qn_p^o D_n}{L_n} \tag{10.13}$$

is the saturation electron current density in the junction. Using hypothesis (9.2) we thus find that

$$I = I_p(4) + I_n(4) = I_p(4) + I_n(2)$$

$$= I_s(e^{q\delta\psi/kT} - 1) \tag{10.14}$$

$$I_s = I_{ps} + I_{ns} \tag{10.15}$$

being the saturation current density.

Formula (10.14) is the fundamental relationship used in rectifier theory. We can of course multiply by the cross section A of the junction and call v the applied bias $\delta\psi$. Then the current $i = IA$ is given by

$$i = AI_s(e^{qv/kT} - 1) = i_s(e^{qv/kT} - 1)\,. \tag{10.16}$$

11. Variation of ϕ_n, ϕ_p and ψ in a Non-equilibrium pnp Transistor

It will be sufficient, as in the last section, to solve for p_n in the continuity and the current density vs. field and concentration gradient equation, to obtain a complete picture. But this time the boundary conditions are slightly different. We shall deduce them by extending the arguments applied to a pn-junction. Figure 8 gives the aspect of a pnp transistor: an n-region (very narrow in reality) is sandwiched between two p-regions. Contacts are made to the 3 regions by an emitter electrode e, a base electrode b and a collector electrode c. We shall again assume that we have equilibrium conditions in 1 and 4, i.e. identity of ϕ_n and ϕ_p. Sometimes this assumption of equilibrium at e and c is expressed by saying that e and c are ohmic constants, meaning that no rectification occurs at these metal-crystal junctions. This implies that ψ is identical on either side and that holes and electrons at e and c have their equilibrium densities.

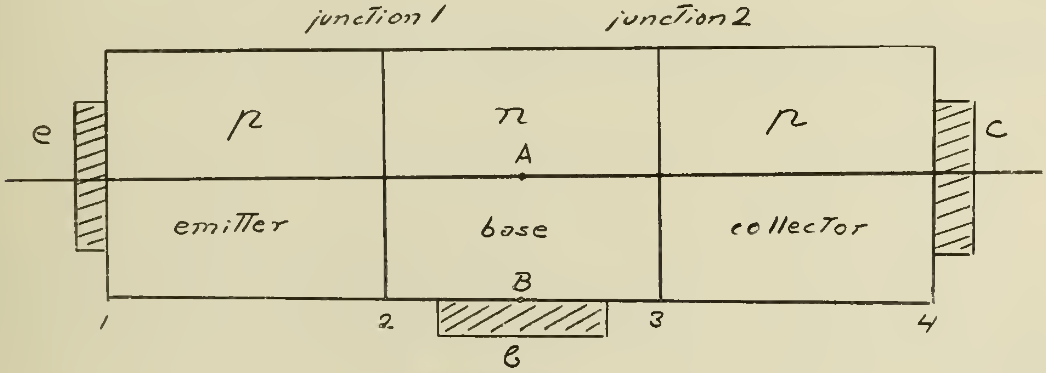

Figure 8

Schematic Diagram of a Transistor

The base contact b is an ohmic contact too, but it is not in direct contact with the interior of the base region and therefore does not imply equilibrium densities on the x-axis of the device. But we shall assume that ψ has the same value in points like A and B having the same x. Taking the base electrode as reference electrode, the applied biasses $\delta\psi_e$ and $\delta\psi_c$ are then simply equal to the potential differences v_e and v_c applied between b and e or c respectively.

By virtue of these hypothesis we can again find the variation of ϕ_n and ϕ_p and ψ in the transistor. We shall use the following principles discussed in section 9:

(1) ϕ_p is constant in the p-regions.

(2) ϕ_n is constant in the n-regions.

(3) At the equilibrium points 1 and 4 $\phi_p = \phi_n$ and the displacement of $\phi_p = \phi_n$ above the $\phi_p^o = \phi_n^o$ line (for equilibrium throughout the transistor) is equal to the applied bias v_e or v_c.

(4) ψ in the n-region (base) has the same value as in the equilibrium case, the base electrode being the reference electrode.

(5) $\phi_n = \phi_n^o$ in the n-region, for ψ is the same as in the equilibrium case and the density $n = n_o$: we can then use (6.5).

As in section 9 we must again distinguish two cases for the collector junction: forward or low reverse bias and high reverse bias. In the latter case ϕ_p and ϕ_n vary considerably in the junction. Figure 9a gives the variation of ϕ_p and ϕ_n for the equilibrium case and for the case of low biasses v_e and v_c. Figure 9b repeats the equilibrium case but also illustrates the case of a strongly reverse biassed collector junction. Luckily the calculations of the next section, based upon Figure 9a, will still be valid when Figure 9b holds, the reason being that the minority carrier density at the collector junction becomes exceedingly small in either picture, i.e. the same boundary conditions hold.

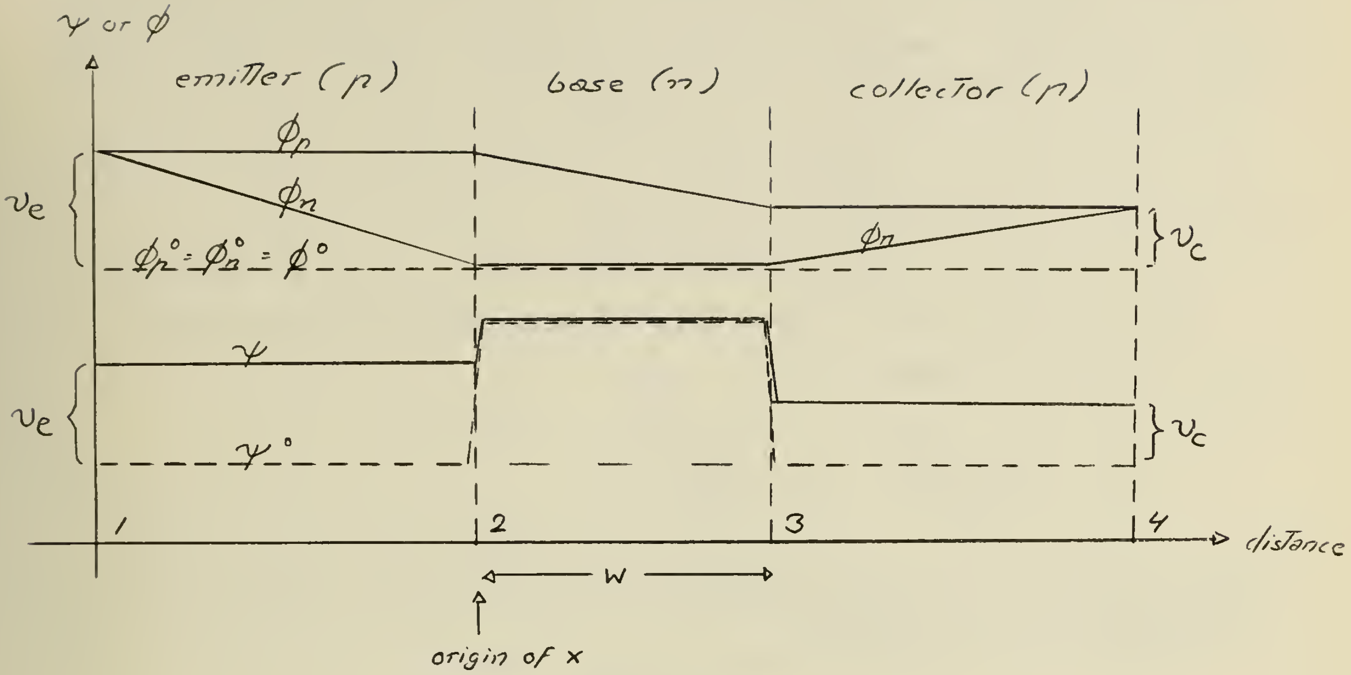
ψ or φ
emitter (p)
base (n)
collector (p)
φp
φn
ve
φp° = φn° = φ°
φn
vc
ψ
ve
ψ°
vc
1
2
3
4
distance
W
origin of x

Figure 9a

ψ, ϕ_p and ϕ_n for a Transistor with Low (Forward) Collector Voltage

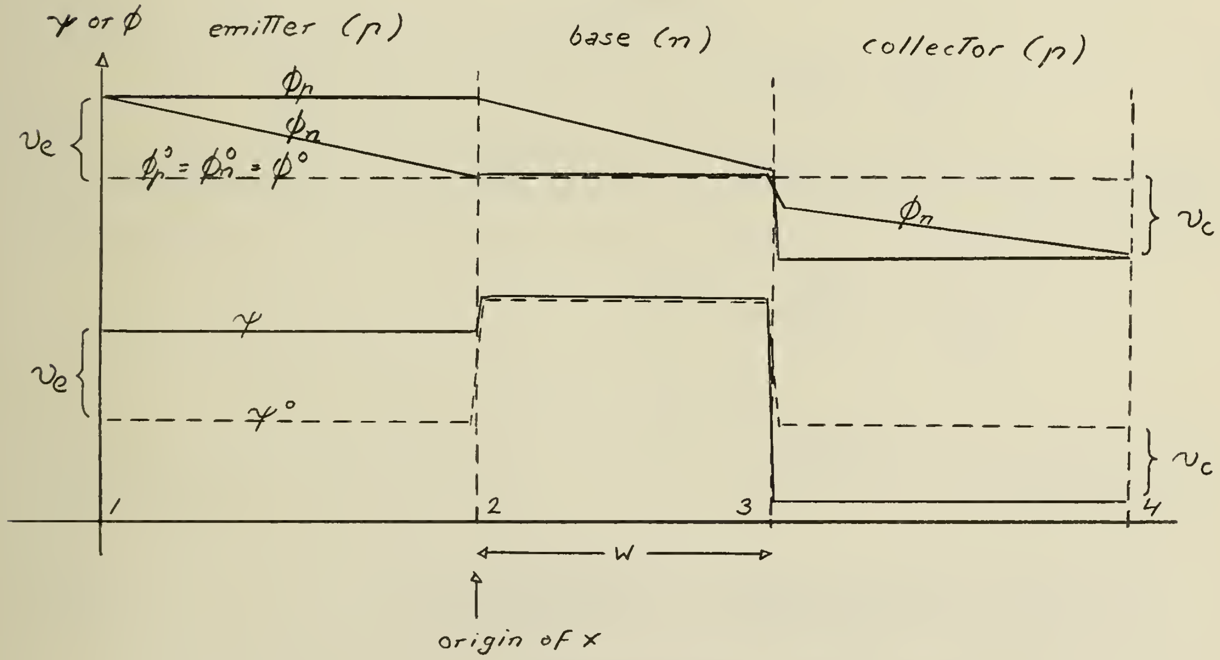
ψ or φ
emitter (p)
base (n)
collector (p)
φp
φn
ve
φp° = φn° = φ°
φn
vc
ψ
ve
ψ°
vc
1
2
3
4
W
origin of x

Figure 9b

ψ, ϕ_n and ϕ_p for a Transistor with High (Reverse) Collector Voltage

12. The Voltage Current Characteristic of a pnp Transistor

As in section 10 we can calculate p(2) and p(3) by taking the difference between ϕ_p and ϕ_n. But $\phi_p(2) - \phi_n(2) = v_e$ and $\phi_p(3) - \phi_n(3) = v_c$. Therefore

$$p_n(2) = p_n^o \, e^{qv_e/kT}$$
$$p_n(3) = p_n^o \, e^{qv_c/kT} \tag{12.1}$$

The solution of the equation (10.6) (applicable because once again $E = 0$ in the base region), i.e.

$$\frac{p_n^o - p_n(x)}{\tau_p} + D_p \frac{\partial^2 p_n(x)}{\partial x^2} = 0 \tag{12.2}$$

which satisfies the boundary conditions (12.1) at points 2 and 3 of coordinates $x = 0$ and $x = w$ is obviously

$$p(x) = p_n^o + \frac{[p_n(3) - p_n^o] \sinh \frac{x}{L_p} + [p_n(2) - p_n^o] \sin \frac{w-x}{L_p}}{\sinh \frac{w}{L_p}} \tag{12.3}$$

where $L_p = \sqrt{D_p \tau_p}$ as before.

This allows us to calculate $I_p(3) = -qD_p \frac{\partial p_n(x)}{\partial x} \Big|_{x = w}$ by (12.3). Introducing

$$I_p^*(v) = \frac{qp_n^o D_p}{L_p} (e^{qv/kT} - 1)$$
$$= I_{ps}(e^{qv/kT} - 1) \tag{12.4}$$

we can then write

$$I_p(3) = \operatorname{csch}\left(\frac{w}{L_p}\right) I_p^*(v_e) - \coth\left(\frac{w}{L_p}\right) I_p^*(v_c). \tag{12.5}$$

Now the total amount of current across the junction in 3, going from left to right, is

$$I(3) = I_p(3) + I_n(3). \tag{12.6}$$

The evaluation of $I_n(3)$ can be done along the same lines as that of $I_p(3)$ by considering the diffusion equation for electrons in the emitter and collector region together with the boundary conditions discussed in section 11. It is, however, easier to observe that the electron flow across each junction is independent of that across the other one. The reason for this is that the majority carrier density in the base, given by (6.5), is constant and equal to that in the equilibrium case ($\psi = \psi_o$ and $\phi_n = \phi_n^o$ in the base by section 11!). Therefore the junction at 3 for example behaves like a pn junction biassed in the reverse direction. Therefore by section 10

$$I_n(3) \cong -I_{ns} = -\frac{qn_p^oD_n}{L_n} \tag{12.7}$$

if we consider the reverse bias big enough.

Calling A the cross section of the two junctions, the current i_c flowing into the collector is $-I(3)A$ for continuity reasons (but $I(3)$ is different from $I(2)$ because of recombination in the base!). We can therefore write

$$i_c = -A\,\mathrm{csch}\left(\frac{w}{L_p}\right)I_p^*(v_e) + A\,\mathrm{coth}\left(\frac{w}{L_p}\right)I_p^*(v_c) + AI_{ns} \tag{12.8}$$

and in a similar way we could establish that

$$i_e = A\,\mathrm{coth}\left(\frac{w}{L_p}\right) I_p^*(v_e) - A\,\mathrm{csch}\left(\frac{w}{L_p}\right) I_p^*(v_c) + AI_{ns} \tag{12.9}$$

This then solves the transistor problem completely, for we can, by differentiation and elimination, calculate the fourpole quantities r_{ik} defined by

$$dv_e = r_{11}di_e + r_{12}di_c$$

$$dv_c = r_{21}di_e + r_{22}di_c$$

and these are the only quantities we must know in the small signal circuit applications.

Remark: As can be seen from the definition of I_{ns} in (12.7) and $I^*_p(v)$ in (12.4), I_{ns} can be neglected if $p_n^o >> n_p^o$. This is usually the case: the emitter and collector regions are more highly doped than the base.

13. Behavior of the Base Region at High Injection Levels

The hypothesis 3 in section 9, i.e. n_o and $p_o >> p_n$ and n_p, led to the conclusion that ϕ_p = constant in a p-region and ϕ_n = constant in an n-region. This fact, together with the nearly constant concentrations p_o and n_o of holes and electrons then allowed us to take ψ as a constant except in transition regions: this means that we neglected electric fields.

We shall now discuss certain features of transistors in which the minority carrier density p in the n-type base is actually much bigger than n_o. This can come about when the density of hole current across the emitter-base junction is high, i.e. for high injection levels. We can then still assume that the variation of ϕ_p and ϕ_n is given by Figure 9a (This means that $p_o >> p_n$, which does not contradict $p_n >> n_o$!), while ψ would now vary in the base region on account of the very variable electron density $n = n_o + p$ (zero space charge!). Thus an electric field appears in the base.

To gain some oversight, let us go back to the general equations (8.5), (8.6), (7.1), (7.2), (5.6), (5.8) and $I = I_p + I_n$:

$$\frac{\partial p}{\partial \tau} = \frac{p^o - p}{\tau_p} - \frac{1}{q}\,\mathrm{div}\, I_p \tag{13.1}$$

$$\frac{\partial n}{\partial \tau} = \frac{n^o - n}{\tau_n} + \frac{1}{q}\,\mathrm{div}\, I_n \tag{13.2}$$

$$I_p = q(\mu_p pE - D_p\,\mathrm{grad}\, p) \tag{13.3}$$

$$I_n = q(\mu_n nE + D_n\,\mathrm{grad}\, n) \tag{13.4}$$

$$I = I_p + I_n \tag{13.5}$$

$$\mathrm{div}\, E = \frac{q}{\epsilon_o \epsilon_r}(p - n + N_d - N_a) \tag{13.6}$$

Let us recall that p^o and n^o are the equilibrium concentrations of carriers (in the p-region $p^o = p_o$ in the n-region $p^o = p^o_n$ etc.) We shall of course retain the zero space charge condition:

$$p - n + N_d - N_a = 0 \tag{13.7}$$

From (13.7) we deduce 3 equations:

$$\mathrm{div}\, E = 0 \tag{13.8}$$

$$\frac{\partial p}{\partial \tau} = \frac{\partial n}{\partial \tau} \tag{13.9}$$

$$\mathrm{grad}\, p = \mathrm{grad}\, n \tag{13.10}$$

Let us now define a dimensionless factor b by

$$\mu_n = b\mu_p \tag{13.11}$$

By Einstein's relationship (see 7.4 and 7.5) we then have

$$D_n = bD_p \tag{13.12}$$

Let us consider an n-region having an electron density of the form discussed in (6.2):

$$n = n_o + p \tag{13.13}$$

(we leave out the subscript n to simplify the formulae). Adding (13.3) and (13.4) we then obtain

$$E = \frac{I - qD_p(b - 1)\,\mathrm{grad}\, p}{q\mu_p[p(b + 1) + bn_o]} \tag{13.14}$$

Therefore, by (13.15)

$$I_p = \frac{pI - qbD_p(2p + n_o)\,\mathrm{grad}\, p}{p(b + 1) + bn_o} \tag{13.15}$$

which means that (13.1) can be written

$$\frac{\partial p}{\partial \tau} + \frac{p - p^o}{\tau_p} = -\,\mathrm{div}\, \frac{pI - qbD_p(2p + n_o)\,\mathrm{grad}\, p}{q[p(b + 1) + bn_o]} \tag{13.16}$$

This, then, is the differential equation for the concentration in an n-region in the most general case, assuming zero space charge. Before integration, one would have to discuss the boundary conditions.

14. Carrier Delay in the Base Region as a Function of Injection Level

Let us consider the mechanism by which signals are transmitted by a transistor. ϕ_p and ϕ_n (for high or low injection level and strong collector bias) are given by Figure 10.

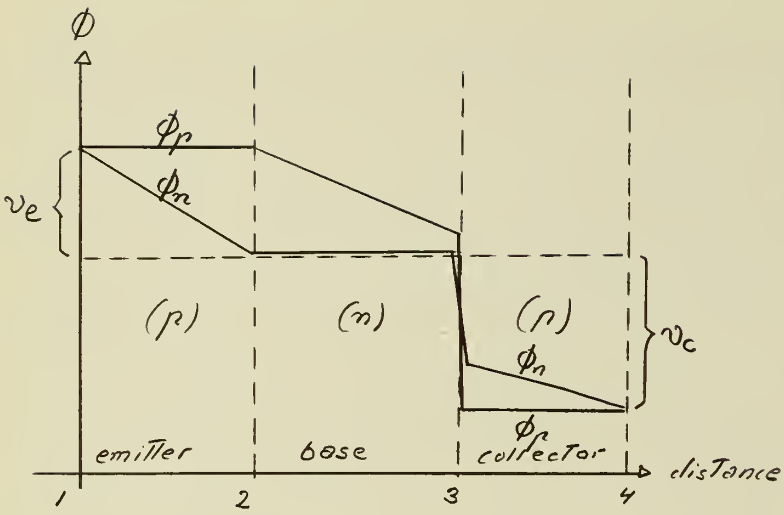

Figure 10

Quasi Fermi-levels in Transistor with High Reverse Collector Voltage

Modifying v_e means--because ϕ_p is nearly horizontal in all cases in the emitter region--modifying the hole concentration at the emitter-base junction: $p_n(2)$ is given by (12.1). This modification is practically instantaneous and there is no signal delay in the emitter region. In an analogous way we could argue that the collector region causes no delay. So we are left with the time of propagation of a perturbation of hole concentration from 2 to 3. This time dealy δ is clearly the base width divided by the average velocity $\bar{\bar{v}}$ of the carriers*:

$$\delta = \frac{w}{\bar{\bar{v}}} \tag{14.1}$$

Now it is clear that in a fixed point we can define an average velocity $\bar{v}$ such that

$$I_p = \bar{v}pq \tag{14.2}$$

In the low injection case there is no electric field in the base and

$$I_p = -qD_p \text{ grad } p \tag{14.3}$$

It follows therefore that

$$\bar{v} = -\frac{D_p}{p} \text{ grad } p$$

i.e.

$$\bar{\bar{v}} = -D_p \overline{\left(\frac{\text{grad } p}{p}\right)}$$

$$\cong -D_p \frac{\overline{\text{grad } p}}{\bar{p}} \tag{14.4}$$

Now for all practical purposed $p_n(3) = 0$ and

---

*) $\bar{\bar{v}}$ indicates an average over all the particles in a given point, averaged over the whole base width.

$$\overline{\text{grad } p} \cong \frac{p_n(3) - p_n(2)}{w} = \frac{-p_n(2)}{w} \tag{14.5}$$

while

$$\bar{p} \cong \lambda p_n(2) \tag{14.6}$$

where λ is a numerical factor of order unity. Thus it follows that

$$\bar{\bar{v}} \cong \frac{D_p}{w}$$

and

$$\delta^1 \cong \lambda \frac{w}{D_p} \tag{14.7}$$

where the upper index 1 refers to the low injection level.

Let us now look at the high injection level case. For this let us go back to equation (13.16), remembering that the square parenthesis is simply $\frac{I_p}{q}$. From (13.4) we obtain, using (13.10), (13.11) and (13.13)

$$E = \frac{I_n - qbD_p \text{ grad } p}{qb\mu_p(n_o + p)} \tag{14.8}$$

which in the high injection level case $p >> n_o$ becomes

$$E = \frac{I_n - qbD_p \text{ grad } p}{qb\mu_p p} \tag{14.9}$$

Substituting this into (13.3) we have

$$I_p = \frac{I_n}{b} - 2qD_p \text{ grad } p \cong - 2qDp \text{ grad } p \tag{14.10}$$

since I_n is negligible by the hypothesis $p_n^o >> n_p^o$.

We see that (14.3) is replaced by

$$I_p = -qD_p' \text{ grad } p \qquad \text{with}$$

$$D_p' = 2D_p \, . \tag{14.11}$$

From there on all calculations remain the same and we obtain

$$\delta^h \cong \lambda \frac{w^2}{2D_p} \tag{14.12}$$

Notice that (14.7), (14.12) and (14.13) are in perfect agreement with the results obtained in section 16 by more sophisticated raasonings.

15. The Current Amplification α of a Transistor

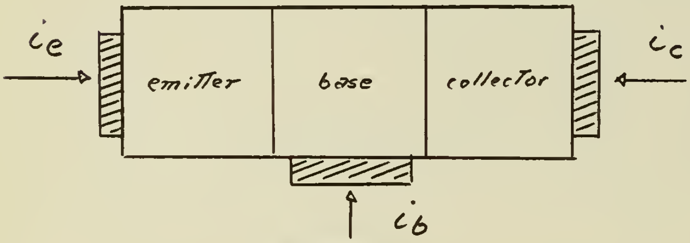

Figure 11

Positive Directions of Currents in a Transistor

In section 12 we found expressions f i_c and i_e as a function of the applied biases v_c and v_e, at least for the low injection level case with zero electric field in the base. Let us call i_b the current flowing from the base electrode into the transistor; then

$$i_e + i_b + i_c = 0 \qquad (15.1)$$

i.e. by using (12.8) and (12.9) with $I_{ns} \sim 0$

$$i_b = -i_e - i_c = \frac{2 \sinh^2\left(\frac{w}{2L_p}\right)}{\sinh\left(\frac{w}{L_p}\right)} \left[I_p^*(v_e) + I_p^*(v_c)\right] \qquad (15.2)$$

which shows clearly that $i_b \neq 0$, or $|i_e| > |i_c|$. This base current clearly corresponds (in the case where most current is due to holes) to recombination of holes and electrons in the base region or flow of electrons across the emitter junction. It is therefore sometimes called "recombination current." Its effect is to decrease the efficiency i_c/i_e of current transfer between emitter and collector.

However it is not very useful to calculate i_c/i_e since in practice we are often interested in the response to small signals. In this case we define the current amplification factor α by

$$\alpha = \left.\frac{\partial i_c}{\partial i_e}\right|_{v_c = \text{constant}} \qquad (15.3)$$

It may seem advantageous to derive α from (12.8) and (12.9) by differentiation. But this would only be valid for very slow variations. Practically we are interested in α at signal frequencies of considerable magnitude. We shall therefore take up the problem from the beginning, i.e. solve equation (13.16) with its time dependent term $\frac{\partial p}{\partial \tau}$ (left out in the steady-state discussion of section 12) and appropriate boundary conditions (v_e variable!)

Let us treat the low injection level case $n_o >> p$ first. Then (13.16) becomes

$$\frac{\partial p}{\partial \tau} + \frac{p - p^o}{\tau_p} = \text{div } D_p \text{ grad } p = D_p \text{ lap } p. \tag{15.4}$$

Introducing as before $L_p = \sqrt{D_p \tau_p}$ and restricting ourselves to the one dimensional case, we have

$$\frac{\partial^2 p}{\partial x^2} - \frac{p - p^o}{L_p^2} = \frac{1}{D_p} \frac{\partial p}{\partial \tau} \tag{15.5}$$

Suppose now that we apply an emitter bias

$$v_e = v_o + v_e^{j\omega t} ,$$

leaving v_c constant according to the definition of α in (15.3). Let $x = 0$ and $x = w$ correspond to the position of the junctions as in section 12. Then the boundary conditions (12.1) are replaced by

$$p_n(2) = p_n^o \, e^{q(v_o + v_e^{j\omega t})/kT}$$

Now assume that $v_o >> v$, then

$$e^{qve^{j\omega t}/kT} \cong 1 + qve^{j\omega t}/kT$$

and

$$p_n(2) = p(0) = p_1 + Pe^{j\omega \tau} \tag{15.6}$$

where

$$p_1 = p_n^o \, e^{qv_o/kT} \tag{15.7}$$

$$P = p_1 qv/kT \tag{15.8}$$

The boundary condition at $x = w$ is the same as the second condition in (12.1) because v_c is constant:

$$p_n(3) = p(w) = p_n^o \, e^{qv_c/kT} \tag{15.9}$$

Notice that in the last equation $v_c < 0$.

The solution of (15.5) satisfying (15.6) and (15.9) is

$$p = p_n^o + \frac{(p_1 - p_n^o)\sinh\left(\frac{w - x}{L_p}\right) + (p(w) - p_n^o)\sinh\left(\frac{x}{L_p}\right)}{\sinh\left(\frac{w}{L_p}\right)}$$

$$+ \frac{Pe^{j\omega\tau}\sinh\left[\sqrt{1+j\omega\tau_p}\left(\frac{w - x}{L_p}\right)\right]}{\sinh\left[\sqrt{1+j\omega\tau_p}\left(\frac{w}{L_p}\right)\right]} \tag{15.10}$$

We can now calculate

$$I_p(w) = -qD_p \left.\frac{\partial p}{\partial x}\right|_{x = w} \tag{15.11}$$

To $I_p(w)$ we must add $I_n(w)$, but again this turns out to be $\sim -I_{ns}$ and therefore negligible if $p_n^o >> n_p^o$. Thus

$$i_c \cong -AI_p(w) \tag{15.12}$$

where A is the cross section of the junctions. But since we are only interested in the alternating part $\tilde{i}_c$ of i_c, it is sufficient to calculate (15.11) for the second term in (15.10). We then obtain

$$\tilde{i}_c = APe^{j\omega\tau}\left(\frac{qD_p\sqrt{1+j\omega\tau_p}}{L_p}\right)\operatorname{csch}\left[\sqrt{1+j\omega\tau_p}\left(\frac{w}{L_p}\right)\right] \tag{15.13}$$

In a similar way we can compute the alternating part $\tilde{i}_e$ of i_e:

$$\tilde{i}_e = APe^{j\omega\tau}\left(\frac{qD_p\sqrt{1+j\omega\tau_p}}{L_p}\right)\coth\left[\sqrt{1+j\omega\tau_p}\left(\frac{w}{L_p}\right)\right] \tag{15.14}$$

and therefore (supposing everywhere v_c = constant)

$$\alpha = \frac{\partial i_c}{\partial i_e} = \frac{\tilde{i}_c}{\tilde{i}_e} = \frac{\operatorname{csch}\left[\sqrt{1+j\omega\tau_p}\left(\frac{w}{L_p}\right)\right]}{\coth\left[\sqrt{1+j\omega\tau_p}\left(\frac{w}{L_p}\right)\right]}$$

i.e.

$$\alpha = \frac{1}{\cosh\left[\sqrt{1 + j\omega\tau_p}\left(\frac{w}{L_p}\right)\right]} \tag{15.15}$$

Remark: For $\omega = 0$ we have

$$\alpha_0 = \frac{1}{\cosh\left(\frac{w}{L_p}\right)} \quad . \tag{15.16}$$

Since w/L_p is quite small in general, $\alpha_o \sim 1$.

16. The α Cutoff Frequency as a Function of the Injection Level

From (15.15) we deduce two facts: first of all there is a phase shift between $\tilde{i}_e$ and $\tilde{i}_c$ at higher frequencies because α is complex. Secondly $|\alpha|$ decreases with ω: the current amplification of a transistor drops at higher frequencies.

It is useful to define a α-cutoff frequency $f_c = \omega_c/2\pi$ such that the corresponding value $|\alpha_c|$ is "3db down." This means that

$$\frac{|\alpha_c|^2}{\alpha_o^2} = \frac{1}{2} \tag{16.1}$$

We shall first examine the low injection case of the preceding section. In order to calculate ω_c we shall suppose that

$$\omega_c\tau_p >> 1. \tag{16.2}$$

Then

$$\sqrt{1 + j\omega_c\tau_p} \simeq \sqrt{j\omega_c\tau_p} = \sqrt{\frac{\omega_c\tau_p}{2}}\,(1+j) \tag{16.3}$$

Let us introduce

$$\eta = \frac{w}{L_p}\sqrt{\frac{\omega_c\tau_p}{2}} = w\sqrt{\frac{\omega_c}{2D_p}} \tag{16.4}$$

by virtue of the definition of the diffusion length L_p. Then (15.15) gives

$$\alpha_c = \frac{1}{\cosh(\eta + \eta j)} = \frac{1}{\cosh\eta\cos\eta + j\,\sinh\eta\sin\eta}$$

and

$$|\alpha_c|^2 = \frac{1}{\cosh^2\eta\cos^2\eta + \sinh^2\eta\sin^2\eta} = \frac{2}{\cos 2\eta + \cosh 2\eta}$$

Using (16.1) this becomes

$$2 = \frac{\cos 2\eta + \cosh 2\eta}{2} \tag{16.5}$$

Solving this transcendental equation we obtain

$$\eta = 1.103$$

which, using (16.4) finally gives

$$\omega_c^1 = 2.434 \frac{D_p}{w^2} \tag{16.6}$$

where the upper index 1 draws attention to the fact that we are talking about the low injection level case.

It seems at a first glance necessary to repeat the calculations of the preceding section for high injection levels. Happily enough this is not necessary; we shall show that everything behaves at high injection levels as if D_p had been replaced by $D_p' = 2D_p$. We only have to repeat the argument of section 14: using

$$I_p \cong -2qD_p \text{ grad } p \tag{16.7}$$

equation (13.16) becomes

$$\frac{\partial p}{\partial \tau} + \frac{p - p^o}{\tau_p} = 2D_p \text{ lap } p$$

$$= D_p' \text{ lap } p \tag{16.8}$$

This proves our contention for the differential equation. The boundary conditions remain unchanged, so all we have to prove is that (15.11) is unchanged if D_p' is substituted for D_p. But this is easy, for we have already shown at the beginning of this section that (13.3)--i.e. the equation for I_p taking account of electric fields--can be written in the form (16.7) in the high injection level case. (15.11) therefore becomes:

$$I_p = -qD'_p \left.\frac{\partial p}{\partial x}\right|_{x = w} \qquad (16.9)$$

All this leads to the conclusion that the reasonings at the beginning of this section are valid if $D_p \rightarrow D'_p = 2D_p$. Calling ω_c^h the value of the α cutoff frequency (divided by 2π) for the high injection level case, we therefore have

$$\omega_c^h = \frac{2.434\, D'_p}{w^2} = 4.868 \frac{D_p}{2} = 2\omega_c^l \qquad (16.10)$$

17. The Transition Region Capacitance

Let us go back to the problem of a single pn junction treated in section 10. But this time we shall take account of the fact that the transition region between points 2 and 4 has a finite width, i.e. of the fact that n and p do not abruptly change at 3 but gradually between 2 and 4. The width will however be taken small enough to allow us to assume that $\phi_p - \phi_n$ in it is constant and equal to $\delta\psi$ (see figure).

We shall simplify the problem by assuming that relative to an origin of x situated between 2 and 4 (but not necessarily in the middle if $n_o \neq p_o$!) we have

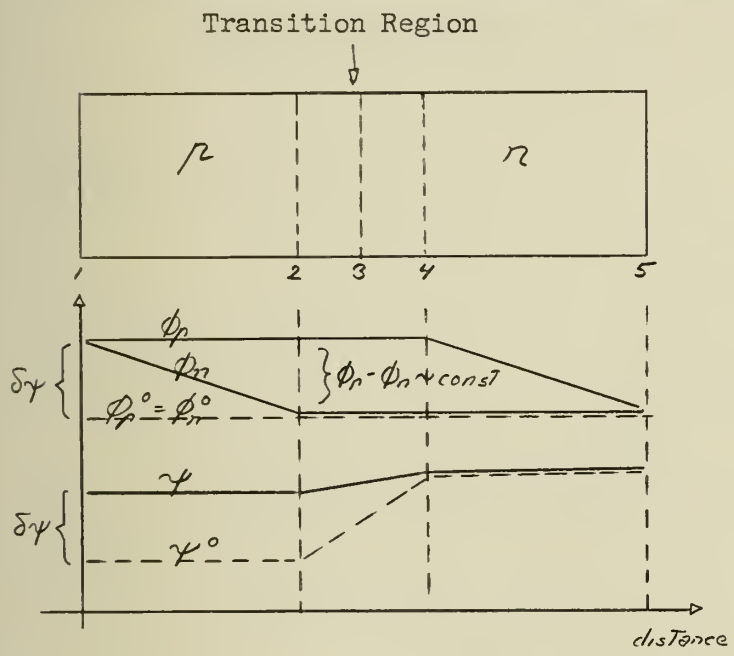

Figure 12
ψ, ϕ_p and ϕ_n in a Junction with Forward Bias

$$n - p = N_d - N_a = ax \; {}^{*)} \qquad (17.1)$$

a and the exact situation of the origin being such that

$$\left.\begin{aligned} n_o &= ax(3) \\ p_o &= ax(2) \end{aligned}\right\} \qquad (17.2)$$

---

*) The first half of the equation is a consequence of space charge neutrality.

Points 2 and 4 will be chosen sufficiently far in the p- and n-regions respectively to warrant the absence of electron flow in 2 and the absence of hole flow in 4. Further we shall neglect recombination between 2 and 4 and as already mentioned, assume space charge neutrality.

Now modifying the applied bias by $d\delta\psi = dv$ modifies ϕ_p, ϕ_n and the hole densities between 2 and 4. Between 1 and 2 $p \sim p_o$, between 4 and 5 $p \sim p_n^o$; i.e. the variation of the total charge of holes Q with a given variation of bias is produced in the transition region alone. The same reasoning leads to the conclusion that the change of the total electronic charge (-Q because of the neutrality condition) is produced in the transition region. This means then that dv produces a drift of charge to the right of dQ in 2 and a drift of electrons to the left (-dQ) in 4: the transition region behaves like a capacitor of capacitance

$$C = \frac{dQ}{dv} \tag{17.3}$$

Using (6.5) we have in the general case

$$\left.\begin{aligned} n &= n_i \; e^{q(\psi-\phi_n)/kT} \\ p &= n_i \; e^{q(\phi_p-\psi)/kT} \end{aligned}\right\} \tag{17.4}$$

Now by figure 12 we have throughout the (narrow) transition region

$$\phi_p - \phi_n = \delta\psi = v. \tag{17.5}$$

Introducing the quantities n_1 and ϕ_1 by

$$\phi_1 = \frac{\phi_p + \phi_n}{2} \tag{17.6}$$

$$n_1 = n_i \; e^{qv/2kT} \tag{17.7}$$

we can then write

$$\left.\begin{aligned} n &= n_1 \; e^{q(\psi-\phi_1)kT} = n_1 e^{u} \\ p &= n_1 \; e^{q(\phi_1-\psi)/kT} = n_1 e^{-u} \end{aligned}\right\} \tag{17.8}$$

where

$$u = q^{(\psi-\phi_1)/kT} \tag{17.9}$$

We can then transform (17.1) into

$$2n_1 \cosh u \, du = a dx \tag{17.10}$$

Let P be the total number of holes between 2 and 4; then

$$P = A\int_2^4 p dx = A\int_{u_2}^{u_4} n_1 e^{-u} \frac{2n_1}{a} \cosh u \, du$$

where A is once again the cross section of the junction and u_4 and u_2 are the values of u at points 2 and 4 respectively. Integrating

$$P = \frac{An_1^2}{a}\int_{u_2}^{u_4} (1 + e^{-2u})du = \frac{An_1^2}{a}\left[(u_4 - u_2) + \frac{e^{-2u_4} - e^{-2u_2}}{2}\right] \tag{17.11}$$

Now at points 2 and 4 we have from (17.8)

$$u_2 = -\ln \frac{p(2)}{n_1} \qquad\qquad u_4 = \ln \frac{n(4)}{n_1} \tag{17.12}$$

(17.11) can thus be written

$$P = \frac{An_1^2}{2a}\left[2(u_4 - u_2) + \left\{\frac{p(2)}{n_1}\right\}^2 - \left\{\frac{n_1}{n(4)}\right\}^2\right]$$

$$\cong \frac{Ap_o^2}{2a} + \frac{An_1^2}{a}(u_4 - u_2) \tag{17.13}$$

where we neglect $\left\{\frac{n_1}{n(4)}\right\}^2 = \frac{n_1^2}{n_o^2}$ because we can assume that the n-region is sufficiently doped to have $n_o >> n_1$; p(2) is replaced p_o. We can now calculate $u_4 - u_2$:

$$u_4 - u_2 = \ln \frac{p(2)n(4)}{n_1^2} = \ln \frac{p(2)n(4)}{n_i^2} + \ln \frac{n_i^2}{n_1^2}$$

$$= \ln \frac{n_o p_o}{n_i^2} - \frac{2qv}{2kT} \tag{17.14}$$

by using (17.7). Replacing n_1 and $u_4 - u_2$ by their value in (17.13)

$$Q = qP = \frac{qAp_o^2}{2a} + \frac{qAn_i^2}{a} \quad e^{qv/kT} \left[\ln \frac{n_o p_o}{n_i^2} - \frac{qv}{kT} \right]$$

i.e.

$$C = \frac{q^2 A n_i^2}{kTa} \left[\ln \frac{n_o p_o}{n_i^2} - \frac{qv}{kT} - 1 \right] \tag{17.15}$$

Let us suppose that the n- and p-regions are strongly doped (n_o and $p_o >> n_i$). Then at room temperature qv/kT is of order unity ($v \sim 30$mv!) and

$$C \cong \frac{q^2 A n_i^2}{kTa} \ln \frac{n_o p_o}{n_i^2} \quad e^{qv/kT} \tag{17.16}$$

Now by (10.14) we have

$$\frac{i + i_s}{i_s} = e^{qv/kT}$$

Calling C_o the value of C for $v = 0$, we obtain

$$C = C_o \frac{i + i_s}{i_s} \quad , \tag{17.17}$$

showing that the transition region capacitance increases with the junction current.

It should be remarked that the reasonings in this section can only be applied to the emitter junction, firstly because we suppose ϕ_p and ϕ_n constant in the transition region (see section 9) and secondly because we assume space charge neutrality. As explained in the next section this latter assumption is not valid for the strongly reverse biased collector junction.

18. The Collector Space Charge Layer

Let us now consider the collector transition region, extending from $-\Delta$ on the left of the junction 3 to $+\Delta$ at the right side of the junction, the exact position of which is given below.

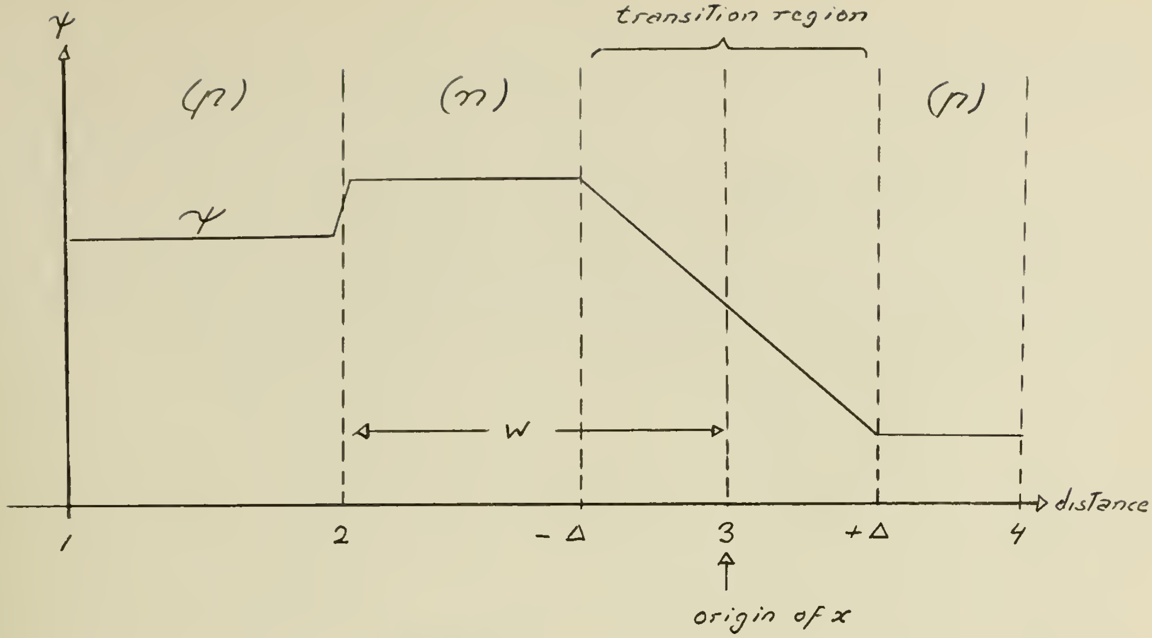

Figure 13

Variation of ψ in a Transistor

We have already explained that for sufficiently high reverse bias the space charge neutrality condition can no longer be satisfied, the reason being that the collector region forms a very good sink for the minority carriers in the base: they are drained out at such a rate that a "depletion layer" is formed in the transition region. In this depletion layer $p \sim 0$ and $n \sim 0$, meaning that

$$\rho = q(N_d - N_a) \tag{18.1}$$

As in section 17 we can then use the second half of (17.1), giving

$$\rho = ax \tag{18.2}$$

which fixes, as we have seen in section 17, the "midpoint" 3 (corresponding to $x = 0$) of the junction.

We can now use (5.8) in the one-dimensional case:

$$\frac{\partial^2 \psi}{\partial x^2} = - \frac{q\,ax}{\epsilon_r \epsilon_o} \tag{18.3}$$

By integrating twice we obtain

$$\psi = - \frac{qax^3}{6\epsilon_r \epsilon_o} + C_1 x + C_2$$

Now it is easy to find C_1:

$$\frac{d\psi}{dx} = -\frac{qax^2}{2\epsilon_r\epsilon_o} + C_1$$

must be approximately zero for $x = \pm\Delta$ when the "normal" p and n regions begin, for ψ is constant in these regions (see section 9). Therefore

$$C_1 = \frac{qa\Delta^2}{2E_r\epsilon_o} \tag{18.4}$$

and if we put

$$u = \psi(+\Delta) - \psi(-\Delta) \tag{18.5}$$

we obtain

$$u = \frac{2aq\Delta^3}{3\epsilon_r\epsilon_o}$$

Now $u = v_c$, therefore

$$\Delta^3 = \frac{3\epsilon_r\epsilon_o v_c}{2aq} \tag{18.6}$$

It is thus evident that the depletion layer extends further and further into the base region as the collector voltage goes up, reducing the base width w to an effective width

$$w_{eff} = w - \Delta . \tag{18.7}$$

When w_{eff} becomes zero we talk about punch-through: the corresponding collector voltage v_c is the punch-through voltage. Above this bias all transistor action is lost.

Further (18.6) and (18.7) show that the transistor equations (with w replaced by w_{eff}) become more complicated, in particular the collector swing in an amplifier modifies w_{eff} constantly and thereby introduces a feedback effect.

Another consequence is a decrease of recombination of injected minority carriers in the base layer since the average carrier diffuses across the narrower base in a shorter time. This corresponds to an increase in α and in the α cutoff frequency.

BOUND BY
HNM

UNIVERSITY OF ILLINOIS-URBANA
3 0112 084228482